

CLAIMS

1. A method for generating an identification signal, comprising the steps of:

accepting as input a monophonic audio signal of limited duration;

translating said monophonic audio signal to a representation of a series of discrete tones; and

producing a control signal from said representation of discrete tones, control signal suitable for causing a transponder to generate a signal,

where said generated signal is human-recognizable as a translation of said monophonic audio signal.

2. A method for generating an identification signal, comprising the steps of:

accepting as input a voice signal of limited duration;

translating said voice signal to a representation of a series of discrete tones; and

producing a control signal from said representation of discrete tones, said control signal suitable for causing a transponder to generate a signal,

where said generated signal is human-recognizable as a translation of said voice signal.

3. The method of claim 2 wherein said generated signal is melodically human-recognizable.

4. The method of claim 2 wherein said generated signal is rhythmically human-recognizable.

5. The method of claim 2 wherein accepting as input further comprises receiving said voice signal over a telephone connection.

6. The method of claim 5 wherein said telephone connection is wireless.

7. The method of claim 2 wherein said step of accepting as input further comprises receiving said voice signal over a microphone attached to a computer.

8. The method of claim 2 wherein said translating step further comprises translating said voice signal to a range of tones within the capability of a mobile telephone audio output synthesizer.

9. The method of claim 2 further comprising the step of transmitting said control signal to a tone-producing output device responsive to said control signal.

10. The method of claim 2 wherein said translating step further comprises the steps of:

generating a digital representation of said voice signal;

dividing said digitized signal into a plurality of frames;

extracting analysis data from each said frame; and

formatting said analysis data into a frame representation.

11. The method of claim 10 wherein said frame representation further comprises a plurality of signal parameters including a time-domain energy measure, a fundamental frequency value, cepstral coefficients, and a cepstral-domain energy measure.

12. The method of claim 11 further comprising the step of determining said time-domain energy measure by multiplying the signal in a selected frame with a mean removed by a window function, summing the square of the result, and normalizing the summed square by the number of samples in said selected frame.

13. The method of claim 12 wherein said window function is a unimodal window function.

14. The method of claim 11 further comprising the step of determining a fundamental frequency of a selected frame by determining the lowest significant periodic component of the signal of said selected frame.

15. The method of claim 11 further comprising the step of determining cepstral coefficients of a selected frame by computing the inverse discrete Fourier transform of the complex natural logarithm of the short-time discrete Fourier transform of the signal of a selected frame, said signal windowed by a window function.

16. The method of claim 11 further comprising the step of determining said cepstral-domain energy measure by determining a short-time cepstral gain with the mean value removed, said short-time cepstral gain normalized by the maximum gain over all frames.

17. The method of claim 11 further comprising the step of determining short-term averages of said plurality of signal parameters.

18. The method of claim 17 further comprising the step of determining each said short-term average over three consecutive frames.

19. The method of claim 17 further comprising the step of determining creating ordinal vectors encoding the number of frames in which directionality of change as determined by said short-term averages remains the same.

20. The method of claim 19 wherein said ordinal vectors further comprise a count of consecutive upward short-term average change in cepstral-domain energy, a count of consecutive downward short-term average change in cepstral-domain energy, a count of consecutive upward short-term average change in fundamental frequency, and a count of consecutive downward short-term average change in fundamental frequency.

21. The method of claim 20 further comprising the step of determining each count for each frame in said signal.

22. The method of claim 10 further comprising the step of segmenting said signal by counting instances of increased signal amplitude in said frames, and

for each instance of increased amplitude, determining a change in each of pitch, energy, and spectral composition in a region around said instance of increased amplitude,

whereby a segment is defined by a start frame having an instance of increased amplitude and an end frame is defined by changes in pitch, energy and spectral composition in relation to selected thresholds.

23. The method of Claim 10 wherein said translating step further comprises grouping said frames into a plurality of regions.

24. The method of claim 23 wherein each said region is determined from a count of consecutive upward short-term average change in cepstral-domain energy followed by a count of consecutive downward short-term average change in cepstral-domain energy.

25. The method of claim 23 further comprising the step of determining the existence of a candidate note start frame in each said region.

26. The method of claim 24 further comprising the step of determining a candidate note start frame in each said region as the last frame within said region in which the count of consecutive upward short-term average change in cepstral-domain energy is not zero.

27. The method of claim 25 further comprising the step of determining which regions of said plurality have a valid note start frame.

28. The method of claim 25, wherein determining a candidate note start frame further comprises the step of determining if the cepstral domain energy of a particular frame is greater than a cepstral domain energy threshold and a frame immediately before said particular frame was below said cepstral domain energy threshold.

29. The method of claim 25, wherein determining a candidate note start frame further comprises the step of determining whether a fundamental frequency range of a particular frame is above a fundamental frequency range threshold and whether an energy range for said particular frame is above an energy range threshold.

30. The method of claim 25, further comprising the step of determining a stop frame corresponding to each start frame.

31. The method of claim 26, further comprising the step of determining a stop frame by locating the first frame after a start frame in which cepstral energy is below said cepstral domain energy threshold.

32. The method of claim 31, further comprising the step of defining the stop frame as a frame between two and ten frames before a subsequent start frame if no frame having cepstral energy below said cepstral domain energy threshold is found.

33. The method of claim 30 further comprising the step of verifying each start and stop frame pair by determining whether a) average voicing probability is above a voicing probability threshold,

b) average short-time energy is above an average short-time energy threshold, and

c) average fundamental frequency is above an average fundamental frequency threshold.

34. The method of claim 30 further comprising the steps of:

forming an initial set of fundamental frequencies from said start and corresponding stop frames;

removing from said initial set those fundamental frequencies having corresponding time-domain energies less than an energy threshold to form a modified set of fundamental frequencies;

removing from said modified set those fundamental frequencies having corresponding voicing probabilities less than a voicing probability threshold to form a twice modified set of fundamental frequencies;

determining a median for each member of said twice modified set;

determining a mode for each member of said twice modified set;

determining a distributional type for each member of said twice modified set with an associated class confidence estimate; and

assigning a MIDI note number to each member of said twice modified set in response to said mode, said median, said distributional type and said class confidence estimate, whereby a note sequence is created.

35. The method of claim 34, further comprising the steps of:

creating a plurality of scales, one for each chromatic pitch class in said note sequence;

assigning a probability to each pitch class, said probability weighted according to scale degree of each note;

comparing each said plurality of scales to said note sequence to find a best fit scale based on occurrences of Tonic, Mediant, and Dominant of a particular scale in comparison to the note sequence; and

selecting the scale with the highest degree of matching.

36. The method of claim 35 wherein said step of assigning probability further comprises:

assigning negative probability weights to the first, sixth, eighth, and tenth scale degrees and positive probability weights to the zeroth, second, fourth, fifth, seventh, and ninth scale degree.

37. The method of claim 36 wherein assigning positive probability further comprises the step of assigning additional positive probability weight to the zeroth, fourth, and seventh scale degree.

38. The method of claim 35 wherein said comparing step further comprises:

ranking said plurality of scales in order of probability; and

comparing said each plurality of scales with said note sequence in order of probability.

39. The method of claim 35, further comprising the steps of:

examining a first pitch pair having a first note having a non-conforming pitch and a second note preceding the first;

if said pitch pair does not conform to voice leading rules, then adjusting said first note unless said adjustment causes dissonance in an adjacent pitch pair.

40. Apparatus for generating an identification signal comprising:

a voice signal receiver;

a translator having as its input a voice signal received by said voice signal receiver and having as its output a representation of discrete tones where an audio presentation of said discrete tones would be human-recognizable as a translation of said voice signal.

41. The apparatus of claim 40 wherein said voice signal receiver comprises an analog telephone receiver.

42. The apparatus of claim 40 wherein said voice signal receiver further comprises a voice-to-digital signal transducer.

43. The apparatus of claim 40 wherein said voice signal receiver further comprises a recording device.

44. The apparatus of claim 40 wherein said translator further comprises a feature estimation module to determine values for at least one time-varying feature of said input signal.

45. The apparatus of claim 44 wherein said translator further comprises a segmentation module responsive to output of said feature estimation module and energy of said input to segment said input signal into notes and a pitch assignment module responsive to signal energy in each segment output by said segmentation module.

46. The apparatus of claim 44 wherein said feature estimation module further comprises a primary feature module, a secondary feature module and a tertiary feature module.

47. The apparatus of claim 46 wherein said primary feature module determines a plurality of values for each of time-domain energy, fundamental frequency, cepstral-domain energy, and voicing probability.

48. The apparatus of claim 46 wherein said secondary feature module determines a plurality of values for each of the secondary features of short-term average change in energy,

short-term average change in fundamental frequency, short-term average change in cepstral coefficient, and short-term average change in cepstral-domain energy.

49. The apparatus of claim 48 wherein each said secondary value is computed over three consecutive frames of said input signal.

50. The apparatus of claim 48 wherein said tertiary feature module determines a plurality of values for at least one of said secondary features.

51. The apparatus of claim 45 wherein said segmentation module further comprises a first-phase segmentation module and a second-phase segmentation module.

52. The apparatus of claim 51 wherein said first-phase segmentation module groups a plurality of successive frames of said input signal into at least one region in response to output of said feature estimation module.

53. The apparatus of claim 52 wherein said region is a plurality of frames in which a change in energy increases immediately followed by frames in which change in energy decreases.

54. The apparatus of claim 53 in which said region has a minimum number of frames.

55. The apparatus of claim 52 wherein said second-phase segmentation module determines if said at least one region has a valid note start frame and if so, determines a stop frame.

56. The apparatus of claim 55 wherein said second-phase segmentation module determines said valid note start frame in response to cepstral domain energy by determining whether a frame has a cepstral domain energy greater than a cepstral domain energy threshold preceded by a frame having a cepstral domain energy less than said cepstral domain threshold.

57. The apparatus of claim 55 wherein said second-phase segmentation module determines a valid note start frame if the fundamental frequency exceeds a fundamental energy threshold and if the non-cepstral domain energy exceeds an energy threshold.

58. The apparatus of claim 52 further comprising a segmentation post-processor to verify said start and stop frame in response to average voicing probability, average short-time energy, and average fundamental frequency of said start and stop frame.

59. The apparatus of claim 45 wherein said pitch assignment module assigns an integer between 32 and 83, said integer corresponding to the MIDI note number for pitch.

60. The apparatus of claim 45 wherein said pitch assignment module comprises an intranote pitch assignment subsystem and an internote pitch assignment subsystem.

61. The apparatus of claim 60 wherein said intranote pitch assignment subsystem determines pitch in response to time-domain energy, voicing probability, median, and mode of each said segment output by said segmentation module.

62. The apparatus of claim 61 wherein said intranote pitch assignment subsystem further comprises an energy thresholding stage to remove from a set of fundamental frequencies for a particular segment those fundamental frequencies whose corresponding time-domain energy are less than an energy threshold to produce a modified set of fundamental frequency for said particular segment.

63. The apparatus of claim 62 wherein said intranote pitch assignment system further comprises a voicing thresholding stage to remove fundamental frequencies from said modified set whose corresponding voicing probabilities are less than a voicing probability threshold to produce a twice-modified set of fundamental frequencies for said particular segment.

64. The apparatus of claim 63 wherein said intranote pitch assignment system further comprises a statistical processing stage to compute a media and a mode for said twice modified fundamental frequency set and to classify said segment as a distributional type in response to said median and said mode.

65. The apparatus of claim 64 wherein said segment is classified as a plurality of distributional types.

66. The apparatus of claim 64 wherein said intranote pitch assignment system further comprises a pitch quantization stage to assign a MIDI note number to said particular segment in response to said median, said mode and said distributional type.

67. The apparatus of claim 66 wherein said statistical processing stage further determines a decision confidence estimate corresponding to the determination of said distributional type, and said pitch quantization stage includes said confidence estimate in the assignment of said MIDI note number.

68. The apparatus of claim 60 wherein said internote pitch assignment subsystem corrects pitches determined by said intranote pitch assignment subsystem.

69. The apparatus of claim 68 wherein said internote pitch assignment subsystem further comprises a key finding stage to assign a scale to a note sequence output by said intranote pitch assignment subsystem.

70. The apparatus of claim 68 wherein said internote pitch assignment subsystem further comprises a pairwise correction stage to examine a pitch and its preceding pitch for conformity to voice-leading rules,

if a pair is determined to be dissonant according to said voice-leading rules, the internote pitch assignment subsystem corrects the pitches of said pair if the pitch adjustment does not cause dissonance in an adjacent pair.